



EDUTECH

Jurnal Teknologi Pendidikan

Journal homepage <https://ejournal.upi.edu/index.php/edutech>



Analisis Kesulitan dan Kesesuaian Butir Soal Teknik Digital Analog Berbasis Model Rasch

Nuridayanti, Awaluddin Tjalla, Iva Sarifah, & Andi Muhammad Ishak

Universitas Negeri Jakarta, Indonesia

*Correspondence: nuridayanti@mhs.unj.ac.id

ABSTRACT	ARTICLE INFO
<p>This study evaluates the quality of the Final Semester Examination instrument for the Digital and Analog Electronics course using content validity and Rasch analysis. The instrument comprised 15 multiple-choice items, validated by four experts and administered to 30 first-semester students. Expert judgment indicated strong content validity in terms of construct relevance, wording clarity, and content appropriateness, with Aiken's V values ranging from 0.813 to 0.938. Rasch analysis showed a stable item difficulty structure, reflected by an item reliability of 0.94 and an item separation index of 4.11, indicating the instrument's capacity to distinguish multiple levels of item difficulty. In contrast, person reliability (0.60) and a standard error of measurement of 2.35 suggest limited discrimination of student ability, likely due to the relatively homogeneous sample and suboptimal alignment between item difficulty and ability distribution. The Wright Map highlights mismatches between item difficulty and student ability, providing diagnostic evidence for improving item distribution and assessment design.</p>	<p>Article History: <i>Submitted/Received 24 Des 2025</i> <i>First Revised 23 Jan 2026</i> <i>Accepted 30 Jan 2026</i> <i>First Available online 05 Feb 2026</i> <i>Publication Date 05 Feb 2026</i></p>
<p>ABSTRAK</p> <p>Penelitian ini mengevaluasi kualitas instrumen Ujian Akhir Semester mata kuliah Teknik Digital dan Analog melalui validitas isi dan analisis Rasch. Instrumen terdiri atas 15 butir soal pilihan ganda yang divalidasi oleh empat pakar dan diujikan kepada 30 mahasiswa semester awal. Hasil validasi ahli menunjukkan validitas isi yang kuat pada aspek relevansi</p>	<p>Keyword: <i>Rasch Model, Item Difficulty, Item Fit, Engineering Education Assessment, Final Examination</i></p>

konstruk, kejelasan redaksi, dan kesesuaian konten, dengan nilai Aiken's V berkisar antara 0,813 – 0,938. Analisis Rasch menunjukkan struktur kesulitan butir yang stabil, ditunjukkan oleh reliabilitas item sebesar 0,94 dan item separation sebesar 4,11, yang mencerminkan kemampuan instrumen membedakan beberapa tingkat kesulitan. Namun, person reliability sebesar 0,60 dan SEM sebesar 2,35 mengindikasikan daya pembeda kemampuan mahasiswa yang masih terbatas, dipengaruhi oleh homogenitas sampel dan ketidaksejajaran tingkat kesulitan butir. *Wright Map* mengungkap ketidaksesuaian antara kesulitan butir dan kemampuan mahasiswa sebagai dasar perbaikan distribusi soal dan desain asesmen.

© 2025 Teknologi Pendidikan UPI

1. PENDAHULUAN

Pengukuran hasil belajar merupakan komponen kunci dalam pendidikan tinggi, terutama pada mata kuliah yang menuntut integrasi pemahaman konseptual dan kemampuan aplikatif. Konteks pendidikan vokasional mekatronika menempatkan mata kuliah Teknik Digital dan Analog sebagai fondasi penting bagi penguasaan logika digital, rangkaian analog, serta aplikasi sistem elektronika dan kontrol. Oleh karena itu, instrumen evaluasi pada mata kuliah ini harus mampu memberikan gambaran kemampuan mahasiswa secara akurat dan adil, karena hasil pengukuran tersebut menjadi dasar pengambilan keputusan akademik dan perbaikan pembelajaran. Kualitas butir soal yang kurang memadai berpotensi menghasilkan interpretasi kemampuan yang keliru dan mengurangi keandalan asesmen.

Kualitas instrumen ujian sangat ditentukan oleh karakteristik butir soal yang digunakan, seperti tingkat kesulitan, daya pembeda, bebas dari bias, serta konsistensi pengukuran terhadap kompetensi yang ditargetkan. Analisis psikometrik diperlukan untuk memastikan instrumen evaluasi tidak hanya valid secara substansial, tetapi juga kuat secara empiris (Wind & Lugu, 2024). Dalam pendidikan teknik, kebutuhan ini semakin krusial karena mahasiswa dihadapkan pada materi yang bersifat hierarkis dan menuntut keterkaitan antara teori dan aplikasi (Saptadi et al., 2025). Ketidaktepatan butir soal dapat berdampak langsung pada ketidakadilan evaluasi dan pengambilan keputusan pedagogis yang tidak tepat.

Seiring berkembangnya pendekatan pengukuran modern, Teori Respons Butir (*Item Response Theory/IRT*) menawarkan kerangka yang lebih kuat dibandingkan teori tes klasik dalam mengevaluasi kualitas instrumen. Salah satu model yang paling luas digunakan adalah *Rasch Model*, yang memetakan kemampuan peserta dan kesulitan butir pada satu skala logit yang sama, sehingga memungkinkan interpretasi pengukuran yang lebih objektif dan bersifat invarian dalam kondisi tertentu (Fischer, 2006; Linacre et al., 2017; Khine, 2020). Selain menghasilkan estimasi kesulitan item, *Rasch Model* menyediakan indikator kesesuaian (*fit statistics*) yang memungkinkan identifikasi butir yang tidak berfungsi sesuai asumsi pengukuran (Planinic et al., 2019; Vincent et al., 2015).

Literatur terbaru dalam penelitian berbasis Rasch menyatakan bahwa stabilitas estimasi parameter item sangat dipengaruhi oleh ukuran sampel, panjang tes, dan kecocokan antara kemampuan peserta dan distribusi kesulitan butir (Feng et al., 2019; Kopp & Jones, 2020; O'Neill et al., 2020). Studi simulasi menunjukkan bahwa meskipun sampel kecil dapat memberikan estimasi awal bagi analisis eksploratori, interpretasi hasil perlu mempertimbangkan konteks, kualitas data, dan stabilitas parameter. Penggunaan Rasch Model dalam pendidikan telah direkomendasikan sebagai pendekatan kuat untuk menilai instrumen yang dirancang dengan struktur kompetensi teknis. Model ini membantu memastikan bahwa evaluasi akademik mencerminkan kompetensi yang kompleks, bukan sekadar akumulasi nilai atau respons kebetulan. Analisis Rasch juga memberikan kontribusi teoritis dan praktis terhadap pengembangan instrumen evaluasi, karena hasilnya dapat digunakan untuk merevisi butir, menentukan bank soal yang terkalibrasi, atau menyusun kembali struktur asesmen (Boone, 2016; Dong et al., 2025; Xue et al., 2025) agar lebih sesuai dengan kurikulum berbasis kompetensi yang diterapkan di pendidikan vokasi.

Evaluasi pembelajaran Teknik Digital dan Analog menjadikan *Rasch Model* relevan karena struktur materi bersifat berjenjang, mulai dari penguasaan konsep dasar hingga integrasi analitik dan aplikatif. Model ini memungkinkan analisis kesesuaian distribusi

tingkat kesulitan butir dengan rentang kemampuan mahasiswa, sekaligus menilai kemampuan instrumen ujian dalam membedakan tingkat kemampuan peserta secara akurat. Selain itu, analisis *item-person map* memberikan gambaran keterpaduan antara tingkat kesulitan soal dan kemampuan mahasiswa, sehingga dapat menilai ketepatan penargetan (*targeting*) instrumen evaluasi.

Meskipun penggunaan *Rasch Model* dalam pengembangan dan validasi instrumen pendidikan telah banyak dilaporkan, sebagian besar penelitian masih berfokus pada pengembangan instrumen baru, asesmen non-kognitif, atau mata pelajaran umum. Penelitian yang secara spesifik menerapkan *Rasch Model* untuk menganalisis kualitas instrumen Ujian Akhir Semester pada mata kuliah teknik dasar di pendidikan vokasi, khususnya Teknik Digital dan Analog, masih relatif terbatas. Selain itu, sedikit studi yang mengaitkan hasil analisis Rasch secara langsung dengan kelayakan instrumen sebagai alat evaluasi kurikulum berbasis kompetensi. Celah inilah yang menjadi landasan kebaruan penelitian ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk: (1) menganalisis struktur kesulitan butir soal UAS mata kuliah Teknik Digital dan Analog menggunakan Rasch Model; (2) mengevaluasi kesesuaian (*fit*) butir dan respon mahasiswa terhadap asumsi model pengukuran; serta (3) menilai kelayakan keseluruhan instrumen ujian dalam merepresentasikan kemampuan mahasiswa secara akurat dan adil.

Hasil penelitian diharapkan dapat memberikan kontribusi empiris terhadap evaluasi kualitas asesmen pada pendidikan vokasi serta menjadi dasar perbaikan instrumen dan pengembangan bank soal yang terkalibrasi.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan tujuan mengevaluasi kualitas psikometrik butir soal Ujian Akhir Semester (UAS) pada mata kuliah Teknik Digital dan Analog. Evaluasi dilakukan melalui dua tahap utama, yaitu validasi ahli secara substansial dan analisis empiris menggunakan Rasch Model untuk butir dikotomus. Pendekatan berurutan ini memungkinkan penilaian instrumen dilakukan baik dari sisi kesesuaian konten maupun kinerja empiris butir soal.

Subjek Penelitian

Subjek penelitian terdiri atas dua kelompok. Kelompok pertama adalah validator ahli yang berjumlah empat orang, terdiri atas dosen pendidikan teknik elektronika, ahli evaluasi pembelajaran, dan ahli bahasa. Para validator bertugas menilai kelayakan butir soal dari aspek konten, konstruksi, dan bahasa. Kelompok kedua adalah mahasiswa yang mengikuti mata kuliah Teknik Digital dan Analog pada semester berjalan, dengan jumlah responden sebanyak 30 orang. Jumlah responden ini dipandang memadai untuk analisis Rasch pada tahap eksploratori, namun sekaligus menjadi keterbatasan penelitian yang perlu dipertimbangkan dalam interpretasi stabilitas parameter item dan generalisasi temuan. Instrumen tes yang dianalisis terdiri atas 15 butir soal pilihan ganda dengan skor dikotomus (benar = 1, salah = 0).

Prosedur Penelitian

Prosedur penelitian dilaksanakan melalui empat tahap. Tahap pertama adalah penyusunan instrumen tes UAS yang disesuaikan dengan Capaian Pembelajaran Mata Kuliah (CPMK) Teknik Digital dan Analog. Setiap butir dirancang untuk mengukur

penguasaan konsep dan pemahaman teknis mahasiswa secara proporsional dan berjenjang.

Tahap kedua adalah validasi ahli. Pada tahap ini, instrumen yang telah disusun dinilai oleh empat validator menggunakan skala penilaian 1 – 5, di mana skor 1 menunjukkan “sangat tidak sesuai” dan skor 5 menunjukkan “sangat sesuai”. Penilaian mencakup tiga aspek utama, yaitu kesesuaian konten dengan CPMK, kejelasan konstruksi butir soal, dan ketepatan penggunaan bahasa.

Tahap ketiga adalah pengumpulan data empiris. Instrumen yang telah dinyatakan layak secara substansial digunakan sebagai soal Ujian Akhir Semester. Respons mahasiswa dikumpulkan dan dikodekan dalam bentuk data dikotomis. Tahap terakhir adalah analisis data menggunakan Rasch Model untuk menilai karakteristik psikometrik butir soal dan kualitas instrumen secara keseluruhan.

Instrumen Pengumpulan Data

Instrumen pertama adalah lembar validasi ahli yang digunakan untuk menilai validitas isi butir soal sebelum digunakan secara empiris. Lembar validasi disusun menggunakan skala penilaian 1 – 5 dan mencakup tiga aspek utama, yaitu kesesuaian konten dengan CPMK, kejelasan konstruksi butir, dan ketepatan penggunaan bahasa.

Hasil penilaian ahli dianalisis menggunakan indeks Aiken’s V (Aiken, 1985) dengan rumus:

$$V = \frac{\sum s}{n(c - 1)} \quad (1)$$

$$s = r - l_0 \quad (2)$$

Dimana:

r = skor penilai,

l_0 = skor terendah pada skala

n = Jumlah validator

c = jumlah kategori penilaian

Nilai V yang mendekati 1 menunjukkan tingkat kesepakatan ahli yang tinggi terhadap relevansi butir. Dalam penelitian ini digunakan ambang $V \geq 0,80$ untuk mengidentifikasi butir dengan validitas isi yang kuat.

Instrumen kedua adalah tes objektif UAS berupa 15 butir soal pilihan ganda yang dikembangkan berdasarkan CPMK mata kuliah Teknik Digital dan Analog. Instrumen yang telah dinyatakan layak secara substansial kemudian digunakan sebagai alat pengumpulan data empiris.

Analisis Data

Analisis data dilakukan menggunakan Rasch Model untuk butir dikotomis dengan bantuan perangkat lunak Winsteps versi 4.0.1. Model dikotomis dipilih karena format soal yang digunakan adalah pilihan ganda dengan dua kategori skor (benar-salah), sehingga sesuai dengan asumsi dasar model Rasch satu parameter.

Persamaan dasar Rasch yang digunakan dalam penelitian ini dinyatakan sebagai berikut:

$$P(X_{ni} = 1) = \frac{e^{(\theta_n - \beta_i)}}{1 + e^{(\theta_n - \beta_i)}} \quad (3)$$

di mana θ_n merepresentasikan kemampuan mahasiswa ke-n dan β_i menunjukkan tingkat kesulitan butir ke-i.

Analisis Rasch mencakup estimasi tingkat kesulitan butir dalam skala logit, statistik kesesuaian item dan person (*infit* dan *outfit mean square*), indeks reliabilitas dan pemisahan (*separation index*), serta interpretasi peta *item-person* (*Wright Map*). *Wright Map* digunakan untuk mengevaluasi keselarasan antara distribusi kemampuan mahasiswa dan tingkat kesulitan butir soal.

Pendekatan kombinasi validasi ahli dan analisis Rasch memungkinkan evaluasi instrumen dilakukan secara substansial dan empiris, sehingga hasil analisis dapat digunakan sebagai dasar objektif untuk menilai kualitas instrumen UAS dan mengidentifikasi kebutuhan revisi butir.

3. HASIL DAN PEMBAHASAN

Hasil Validasi Ahli

Hasil validasi yang dilakukan oleh empat pakar terhadap 15 butir instrumen menunjukkan bahwa seluruh butir memiliki indeks Aiken's V di atas 0.80, sebagaimana ditampilkan pada **Tabel 1**. Nilai tersebut menegaskan bahwa adanya kesepakatan kuat di antara para ahli mengenai kesesuaian konstruk, kejelasan redaksi, dan relevansi konten, sehingga instrumen dinilai layak secara validitas isi untuk dianalisis lebih lanjut menggunakan model Rasch. Temuan ini sejalan dengan praktik pengembangan instrumen yang menekankan bahwa validitas isi merupakan prasyarat penting sebelum analisis psikometrik dilakukan (Clark & Watson, 2019; Siregar, 2025; Tajuddin et al., 2025; Zamanzadeh et al., 2015).

Meskipun seluruh nilai Aiken's V berada pada kategori valid, para ahli tetap memberikan sejumlah masukan untuk penyempurnaan teknis dan redaksional. Saran tersebut mencakup perlunya memperjelas parameter analisis pada soal rangkaian, meningkatkan ketepatan teknis distraktor agar mencerminkan miskonsepsi umum di laboratorium, serta memastikan konsistensi notasi elektronika. Selain itu, pakar menekankan pentingnya diagram rangkaian atau timing diagram pada soal digital serta penyesuaian istilah teknis agar tidak menimbulkan ambiguitas. Pada aspek bahasa, beberapa butir disarankan dipadatkan dan distandarkan gaya bahasanya. Walaupun demikian, seluruh masukan tersebut bersifat penyempurnaan dan tidak mengurangi penilaian validitas isi secara kuantitatif, sehingga instrumen tetap dinilai baik dan siap dianalisis pada tahap berikutnya.

Table 1. Hasil analisis validitas isi

No. Butir	Skor Pakar				Indeks Aiken's <i>v</i>	Ket.
	A	B	C	D		
1	4	4	5	4	0.813	Valid
2	4	5	4	4	0.813	Valid
3	4	4	5	4	0.813	Valid

No. Butir	Skor Pakar				Indeks Aiken's v	Ket.
	A	B	C	D		
4	5	5	5	4	0.938	Valid
5	5	5	5	4	0.938	Valid
6	4	5	5	3	0.813	Valid
7	5	5	5	4	0.938	Valid
8	4	5	5	4	0.875	Valid
9	5	5	5	4	0.938	Valid
10	5	5	5	4	0.938	Valid
11	5	5	5	4	0.938	Valid
12	5	5	5	3	0.875	Valid
13	4	5	5	4	0.875	Valid
14	4	5	5	3	0.813	Valid
15	4	5	5	3	0.813	Valid

Gambaran Umum Kualitas Pengukuran Berdasarkan Model Rasch

Analisis menggunakan model Rasch memberikan landasan pengukuran yang lebih kokoh dan informatif dalam mengevaluasi kualitas butir soal Ujian Akhir Mata Kuliah Teknik Digital dan Analog dibandingkan pendekatan teori tes klasik. Model ini tidak hanya memeriksa reliabilitas dan validitas instrumen secara agregat, tetapi juga memungkinkan penelusuran hubungan struktural antara kesulitan butir dan kemampuan responden secara simultan dan invarian. Nilai item reliability sebesar 0,94 dengan item separation sebesar 4,11 pada **Gambar 1** mengindikasikan bahwa hierarki kesulitan butir tersusun secara stabil dan mampu membedakan beberapa strata kesulitan secara bermakna. Dalam kerangka Rasch, separation di atas 4 menunjukkan bahwa butir dapat dikelompokkan ke dalam setidaknya empat hingga lima tingkat kesulitan yang berbeda secara konsisten (Bond et al., 2020; Fitrah et al., 2024). Dengan demikian, instrumen ini memiliki presisi tinggi dalam mengukur variasi kesulitan konten Teknik Digital dan Analog serta memenuhi prinsip invariansi pengukuran, sehingga hasil estimasi kesulitan butir relatif bebas dari karakteristik sampel.

SUMMARY OF 30 MEASURED PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	16.5	15.0	-4.80	.48	.72	-.3	.97	.0
P.SD	3.3	.0	.75	.02	.22	.5	.39	1.0
S.SD	3.4	.0	.77	.02	.22	.5	.40	1.1
MAX.	26.0	15.0	-2.77	.50	1.37	.8	2.20	3.1
MIN.	12.0	15.0	-5.87	.37	.42	-1.3	.50	-1.9
REAL RMSE	.48	TRUE SD	.58	SEPARATION	1.19	PERSON	RELIABILITY	.59
MODEL RMSE	.48	TRUE SD	.58	SEPARATION	1.21	PERSON	RELIABILITY	.60
S.E. OF PERSON MEAN = .14								

PERSON RAW SCORE-TO-MEASURE CORRELATION = 1.00

CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .51 SEM = 2.35

SUMMARY OF 15 MEASURED ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	31.0	30.0	.00	.40	.94	.0	.97	.0
P.SD	56.3	.0	1.76	.07	.27	1.2	.28	1.3
S.SD	58.2	.0	1.82	.07	.28	1.3	.29	1.3
MAX.	248.0	30.0	2.19	.48	1.20	1.7	1.34	2.0
MIN.	6.0	30.0	-5.61	.15	.16	-3.9	.16	-3.9
REAL RMSE	.42	TRUE SD	1.71	SEPARATION	4.11	ITEM	RELIABILITY	.94
MODEL RMSE	.41	TRUE SD	1.72	SEPARATION	4.24	ITEM	RELIABILITY	.95
S.E. OF ITEM MEAN = .46								

Gambar 1. *Person and item reliability*

Sebaliknya, rendahnya nilai *person reliability* sebesar 0,60 dan *person separation* sebesar 1,72 menunjukkan bahwa instrumen belum optimal dalam membedakan tingkat kemampuan mahasiswa (Van Zile-Tamsen, 2017). Temuan ini tidak serta merta mengindikasikan kelemahan butir soal, melainkan lebih merefleksikan keterbatasan variasi kemampuan responden serta ketidaksesuaian antara distribusi kesulitan butir dan kesiapan kognitif mahasiswa semester awal.

Reliabilitas person pada Rasch merepresentasikan kemampuan instrumen dalam memisahkan individu ke dalam strata kemampuan yang bermakna berdasarkan estimasi logit, dan sangat bergantung pada heterogenitas kemampuan sampel serta keselarasan antara kemampuan responden dan tingkat kesulitan item (Bond & Fox, 2015). Dengan jumlah responden yang relatif kecil ($n = 30$) dan karakteristik mahasiswa semester awal yang cenderung homogen, rentang kemampuan menjadi terbatas (*restricted range*), sehingga kemampuan instrumen untuk membedakan person secara tajam juga menurun.

Nilai estimasi *Standard Error of Measurement* (SEM) sebesar 2,35 memberikan informasi mengenai presisi pengukuran dan memungkinkan interpretasi skor mahasiswa secara lebih bertanggung jawab dengan mempertimbangkan ketidakpastian estimasi (AERA, 2014). Dengan demikian, meskipun *person reliability* rendah, instrumen Ujian Akhir Semester tetap berfungsi secara konseptual dan psikometrik sebagai alat

evaluasi yang layak, sekaligus menyediakan dasar empiris yang kuat bagi dosen untuk merevisi distribusi kesulitan butir dan meningkatkan daya diskriminasi instrumen pada implementasi selanjutnya.

Struktur Kesulitan Butir dan Kesesuaian Model

Hasil analisis item fit pada **Gambar 2** menunjukkan bahwa sebagian besar butir berada dalam rentang nilai *infit* dan *outfit* Mean Square (MNSQ) yang dapat diterima, yaitu sekitar 0,7-1,3. Hal ini menunjukkan bahwa respons mahasiswa secara umum konsisten dengan prediksi model Rasch. Namun, beberapa butir, khususnya BUTI09, BUTI10, BUTI11, BUTI13, dan BUTI14 yang menunjukkan nilai outfit MNSQ di atas 1,20 dan korelasi *Point-Measure Correlation* (PTMEA) yang relatif rendah, yang mengindikasikan adanya pola respons yang tidak sepenuhnya sesuai dengan model.

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL		INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		ITEM
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
9	16	30	.48	.39	1.16	1.4	1.34	2.0	A	.08	.33	60.0	63.0	BUTI09
10	17	30	.33	.39	1.20	1.7	1.25	1.4	B	.05	.32	56.7	64.0	BUTI10
11	10	30	1.41	.41	1.14	.9	1.24	1.1	C	.13	.34	66.7	70.7	BUTI11
13	23	30	-.68	.45	1.12	.4	1.17	.5	D	.08	.26	76.7	76.8	BUTI13
14	7	30	1.97	.46	1.14	.6	1.16	.6	E	.15	.32	80.0	77.7	BUTI14
1	24	30	-.88	.46	1.02	.2	1.14	.4	F	.13	.24	80.0	80.0	BUTI01
8	13	30	.93	.39	1.06	.5	1.14	.9	G	.23	.34	70.0	65.4	BUTI08
7	18	30	.18	.39	1.04	.4	.97	-.1	H	.28	.31	53.3	64.9	BUTI07
5	15	30	.63	.39	1.03	.3	.98	-.1	h	.32	.33	56.7	63.0	BUTI05
6	16	30	.48	.39	.99	.0	.96	-.2	g	.35	.33	66.7	63.0	BUTI06
4	19	30	.02	.40	.96	-.2	.90	-.4	f	.37	.30	70.0	66.4	BUTI04
12	6	30	2.19	.48	.91	-.2	.87	-.2	e	.41	.31	83.3	80.7	BUTI12
2	25	30	-1.10	.46	.83	.1	.80	-.1	d	.29	.24	83.3	83.0	BUTI02
15	12	30	1.09	.40	.83	-1.2	.77	-1.4	c	.57	.34	73.3	67.0	BUTI15
3	27	30	-1.47	.38	.44	-.2	.64	-.1	b	.26	.27	86.7	84.1	BUTI03
	248	30	-5.61	.15	.16	-3.9	.16	-3.9	a	1.00	.70	66.7	31.2	JUMLAH
MEAN	31.0	30.0	.00	.40	.94	.0	.97	.0				70.6	68.8	
P. SD	56.3	.0	1.76	.07	.27	1.2	.28	1.3				10.2	12.3	

Gambar 2. Analisis item fit

Fenomena *misfit* pada beberapa butir dan responden dapat dipahami sebagai tuntutan kognitif yang relatif tinggi pada sebagian besar soal. Butir yang menuntut integrasi konsep abstrak dan penalaran tingkat analisis dan sintesis cenderung menghasilkan pola respons yang tidak stabil ketika diberikan kepada mahasiswa yang masih berada pada tahap awal pembentukan skema konseptual (Boone, 2016; Hsu et al., 2024; Maric et al., 2023). Dengan demikian, *misfit* yang muncul lebih tepat diinterpretasikan sebagai konsekuensi dari ketidaksejajaran tingkat kesulitan instrumen dengan perkembangan belajar mahasiswa, bukan sebagai kegagalan konstruk pengukuran.

Tabel 2 menyajikan ringkasan hasil analisis *item fit* berdasarkan model Rasch untuk mengevaluasi kualitas psikometrik setiap butir instrumen. Evaluasi dilakukan dengan mempertimbangkan nilai *Infit* dan *Outfit* MNSQ sebagai indikator kesesuaian butir terhadap model, serta PTMEA sebagai ukuran daya diskriminasi butir terhadap konstruk yang diukur. Kombinasi ketiga indikator tersebut digunakan untuk menilai apakah setiap butir berfungsi secara optimal, menunjukkan kecenderungan *overfit* atau *misfit*, serta

untuk menentukan status butir, apakah dipertahankan atau direvisi guna meningkatkan kualitas pengukuran secara keseluruhan.

Tabel 2. Ringkasan Kesesuaian Butir (*Item Fit*), Daya Diskriminasi, dan Keputusan Revisi Instrumen

Item	Infit MNSQ	Outfit MNSQ	PTMEA-Corr	Interpretasi	Status	Implikasi Revisi
Buti01	1.02	1.14	0.13	Fit, sedikit noise	Dipertahankan	-
Buti02	0.83	0.80	0.29	Overfit ringan, diskriminasi baik	Dipertahankan	-
Buti03	0.44	0.64	0.26	Overfit, item sangat mudah	Revisi	Perbaiki distraktor agar tidak terlalu jelas
Buti04	0.96	0.90	0.37	Fit sangat baik	Dipertahankan	-
Buti05	1.03	0.98	0.32	Fit baik	Dipertahankan	-
Buti06	0.99	0.96	0.35	Fit sangat baik	Dipertahankan	-
Buti07	1.04	0.97	0.28	Fit stabil	Dipertahankan	-
Buti08	1.06	1.14	0.23	Masih fit, sedikit ketidakterdugaan	Dipertahankan	-
Buti09	1.16	1.34	0.08	Misfit signifikan, diskriminasi rendah	Revisi	Perjelas stimulus soal untuk menghindari interpretasi ganda
				Misfit signifikan, distraktor lemah	Revisi	Perbaiki kualitas distraktor agar berfungsi secara merata dan mengurangi tebakan acak
Buti10	1.20	1.25	0.05			Selaraskan level kesukaran dengan kemampuan target mahasiswa
Buti11	1.14	1.24	0.13	Misfit ringan	Revisi	
Buti12	0.91	0.87	0.41	Fit baik	Dipertahankan	-
				Diskriminasi lemah	Revisi	Revisi struktur opsi jawaban untuk meningkatkan daya beda antar tingkat kemampuan
Buti13	1.12	1.17	0.08			Klarifikasi konteks teknis dan eliminasi informasi yang tidak relevan
Buti14	1.14	1.16	0.15	Misfit ringan, noise konten	Revisi	
Buti15	0.83	0.77	0.57	Overfit, diskriminasi sangat kuat	Dipertahankan	-

Analisis Person Fit

Hasil analisis *person fit* pada **Gambar 3** menunjukkan bahwa sebagian besar mahasiswa memberikan pola respons yang relatif konsisten dengan model Rasch. Namun demikian, beberapa responden, seperti Resp18 dan Resp01, menunjukkan nilai *outfit* MNSQ di atas 1,50. Pola ini mencerminkan adanya respons tidak stabil, seperti menjawab benar pada butir sulit tetapi salah pada butir yang relatif lebih mudah.

Person misfit dalam literatur Rasch, diinterpretasikan sebagai indikasi adanya noise dalam proses pengukuran, yang dapat bersumber dari faktor non-kognitif seperti menebak, kelelahan, kurangnya fokus, atau miskonsepsi terhadap soal (Linacre et al.,

2017). Namun, dalam penelitian ini, pola *person misfit* tidak dapat dilepaskan dari karakteristik responden sebagai mahasiswa pada semester semester awal. Penelitian dalam pendidikan teknik menunjukkan bahwa pada fase ini, pemahaman mahasiswa sering kali masih bersifat parsial dan terfragmentasi, sehingga respons terhadap asesmen berbasis pemahaman mendalam menjadi tidak konsisten (Mayer, 2020). Kondisi ini menjelaskan mengapa jawaban benar pada butir tertentu tidak selalu mencerminkan penguasaan konsep yang stabil, melainkan bisa muncul secara sporadis.

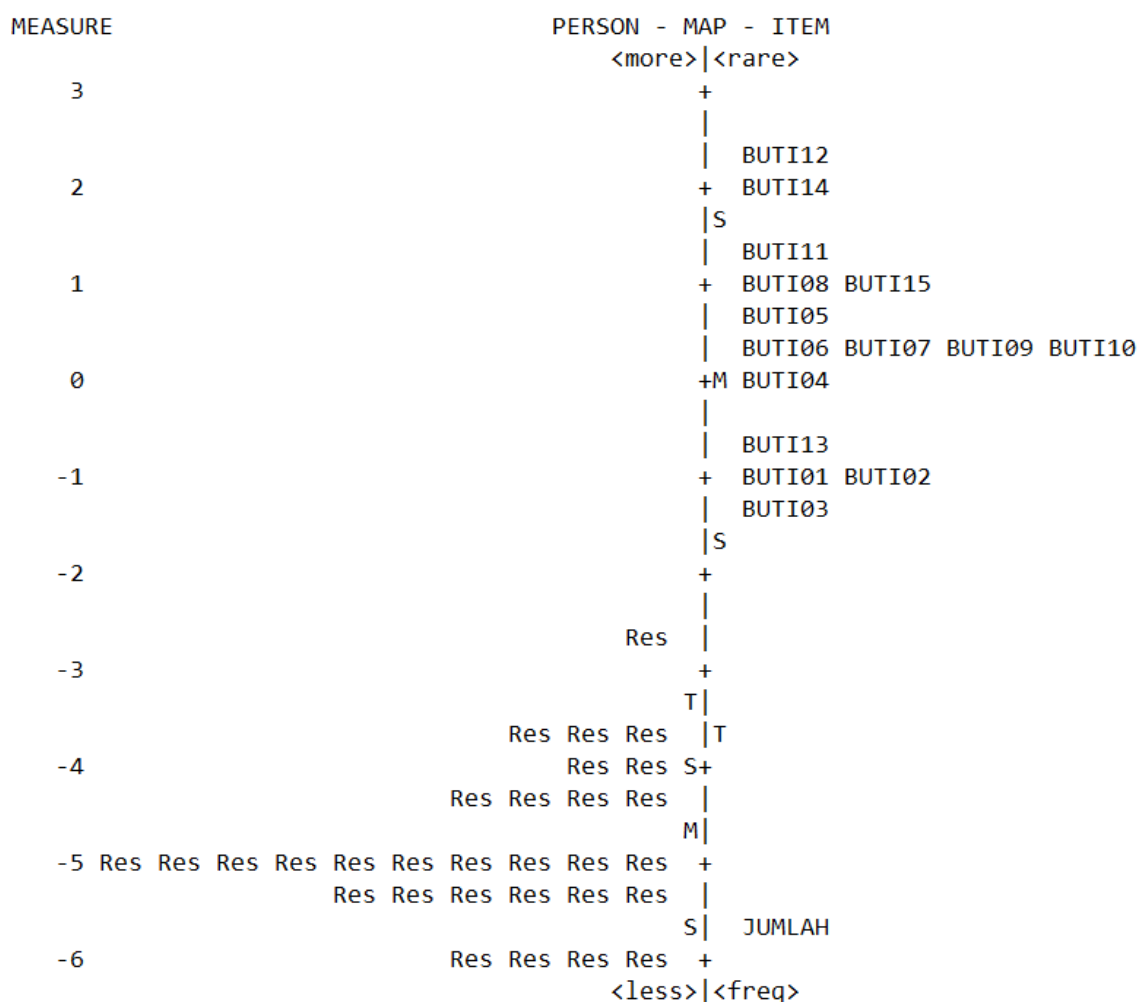
PERSON STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	TOTAL MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	PERSON	
18	16	16	-4.90	.48	1.37	.8	2.20	3.1	A	.73	.85	43.8	66.6	Resp18
1	14	16	-5.38	.50	1.18	.5	1.82	2.2	B	.77	.85	50.0	67.0	Resp01
3	20	16	-4.02	.46	.70	-.4	1.56	.9	C	.83	.84	75.0	71.3	Resp03
22	18	16	-4.44	.47	1.02	.2	1.52	1.3	D	.79	.84	50.0	67.6	Resp22
24	12	16	-5.87	.49	1.14	.5	1.45	1.1	E	.80	.84	56.3	70.8	Resp24
20	18	16	-4.44	.47	.78	-.2	1.14	.5	F	.84	.84	75.0	67.6	Resp20
4	26	16	-2.77	.37	.48	-.1	1.12	.5	G	.82	.83	75.0	75.7	Resp04
17	14	16	-5.38	.50	.83	-.2	1.09	.4	H	.85	.85	75.0	67.0	Resp17
7	16	16	-4.90	.48	.76	-.3	1.04	.2	I	.85	.85	68.8	66.6	Resp07
16	14	16	-5.38	.50	.91	.0	1.04	.2	J	.84	.85	62.5	67.0	Resp16
19	16	16	-4.90	.48	.78	-.3	1.04	.2	K	.84	.85	68.8	66.6	Resp19
23	12	16	-5.87	.49	.84	.2	1.02	.2	L	.86	.84	68.8	70.8	Resp23
30	20	16	-4.02	.46	.72	-.3	.93	.1	M	.84	.84	62.5	71.3	Resp30
25	22	16	-3.58	.47	.66	-.3	.92	.3	N	.85	.83	68.8	75.0	Resp25
5	14	16	-5.38	.50	.75	-.4	.90	-.2	O	.87	.85	75.0	67.0	Resp05
6	16	16	-4.90	.48	.70	-.5	.86	-.4	o	.86	.85	68.8	66.6	Resp06
2	16	16	-4.90	.48	.66	-.6	.82	-.5	n	.87	.85	68.8	66.6	Resp02
12	16	16	-4.90	.48	.61	-.7	.79	-.6	m	.87	.85	81.3	66.6	Resp12
11	18	16	-4.44	.47	.60	-.7	.76	-.6	l	.87	.84	75.0	67.6	Resp11
15	12	16	-5.87	.49	.75	.1	.73	-.6	k	.89	.84	68.8	70.8	Resp15
10	18	16	-4.44	.47	.57	-.7	.70	-.8	j	.87	.84	75.0	67.6	Resp10
27	16	16	-4.90	.48	.60	-.7	.70	-1.0	i	.88	.85	68.8	66.6	Resp27
14	14	16	-5.38	.50	.60	-.8	.69	-1.0	h	.89	.85	87.5	67.0	Resp14
29	16	16	-4.90	.48	.58	-.8	.68	-1.1	g	.88	.85	68.8	66.6	Resp29
28	16	16	-4.90	.48	.56	-.8	.66	-1.2	f	.89	.85	68.8	66.6	Resp28
8	12	16	-5.87	.49	.65	.0	.63	-.9	e	.91	.84	81.3	70.8	Resp08
21	22	16	-3.58	.47	.49	-.6	.64	.0	d	.87	.83	81.3	75.0	Resp21
9	14	16	-5.38	.50	.57	-.8	.61	-1.4	c	.90	.85	75.0	67.0	Resp09
13	22	16	-3.58	.47	.44	-.7	.50	-.1	b	.88	.83	81.3	75.0	Resp13
26	16	16	-4.90	.48	.42	-1.3	.50	-1.9	a	.91	.85	93.8	66.6	Resp26
MEAN	16.5	16.0	-4.80	.48	.72	-.3	.97	.0				70.6	68.8	
P.SD	3.3	.0	.75	.02	.22	.5	.39	1.0				10.6	3.0	

Gambar 3. Person statistic

Wright Map

Wright Map pada **Gambar 4** menunjukkan bahwa distribusi kemampuan mahasiswa terkonsentrasi pada rentang -3 hingga -6 logit, sedangkan tingkat kesulitan butir berada pada rentang 0 hingga +3 logit. Pola ini mengindikasikan ketidaksesuaian antara distribusi kemampuan responden dan tingkat kesulitan instrumen, yang dalam kerangka Rasch mencerminkan masalah penargetan instrumen terhadap populasi yang diuji.



Gambar 4. *Wright Map*

Dalam pengukuran Rasch, kondisi optimal dicapai ketika rerata kemampuan responden relatif sejajar dengan rerata kesulitan item, sehingga informasi pengukuran tersebar merata di sepanjang kontinum kemampuan (Bond et al., 2020). Ketidaksejajaran yang tampak pada **Gambar 4** menyebabkan sebagian besar mahasiswa berada di bawah tingkat kesulitan item yang tersedia, sehingga instrumen hanya memberikan informasi terbatas pada rentang kemampuan rendah. Situasi ini memicu *floor effect*, yaitu kondisi ketika mayoritas responden memperoleh skor rendah karena tidak mampu menjawab sebagian besar butir, termasuk butir yang relatif paling mudah.

Konsekuensi langsung dari kondisi tersebut tercermin pada rendahnya *person separation* dan *person reliability*. Karena hampir seluruh butir berada di atas tingkat kemampuan mahasiswa, instrumen kesulitan membedakan secara presisi antara mahasiswa berkemampuan sangat rendah dan yang sedikit lebih tinggi. Temuan ini selaras dengan studi psikometrik berbasis Rasch yang menunjukkan bahwa ketiadaan butir pada rentang kemampuan rendah menyebabkan menurunnya daya diskriminasi dan presisi estimasi kemampuan responden (Xue et al., 2025).

Implikasi pedagogis dari temuan ini terutama relevan bagi asesmen pada mahasiswa semester awal. Instrumen evaluasi pada tahap ini perlu dirancang dengan sebaran tingkat kesulitan yang lebih luas, mencakup butir sangat mudah hingga butir dengan

kesulitan tinggi. Strategi tersebut memungkinkan instrumen menangkap variasi kemampuan mahasiswa secara lebih sensitif, sekaligus meningkatkan kualitas pemetaan kemampuan dan ketepatan pengukuran pada fase awal perkuliahan.

Hasil analisis ini membantu dosen dalam menyeimbangkan tingkat kesulitan soal, mengidentifikasi serta memperbaiki butir yang kurang berfungsi secara optimal, dan merancang instrumen evaluasi selanjutnya agar lebih selaras dengan capaian pembelajaran mata kuliah Teknik Digital dan Analog serta karakteristik kemampuan mahasiswa yang lebih heterogen. Selain itu, temuan ini menegaskan pentingnya penerapan pembelajaran bertahap yang didukung oleh asesmen formatif pada tahap awal perkuliahan, sehingga kesiapan belajar mahasiswa dapat dipetakan secara lebih akurat sebelum penerapan asesmen sumatif dengan tuntutan kognitif yang lebih tinggi.

4. SIMPULAN

Penelitian ini menunjukkan bahwa instrumen Ujian Akhir Semester mata kuliah Teknik Digital dan Analog memiliki validitas isi yang kuat dan kualitas psikometrik item yang memadai berdasarkan analisis Rasch. Hasil validasi ahli dan estimasi Rasch mengonfirmasi struktur kesulitan item yang stabil, nilai item reliability dan item separation yang tinggi, serta terpenuhinya prinsip invariansi pengukuran. Temuan ini menegaskan bahwa instrumen layak digunakan untuk mengevaluasi capaian pembelajaran kognitif pada mata kuliah tersebut.

Namun, rendahnya *person reliability* dan *person separation* menunjukkan keterbatasan instrumen dalam membedakan kemampuan mahasiswa semester awal. Kondisi ini terutama disebabkan oleh keterbatasan desain uji coba, meliputi jumlah butir yang terbatas, ukuran sampel yang kecil, dan homogenitas kemampuan responden. Temuan ini mencerminkan ketidaksejajaran antara tingkat kesulitan soal dan kemampuan mahasiswa, bukan kelemahan konstruk instrumen.

Secara praktis, hasil penelitian ini menekankan perlunya penyesuaian distribusi tingkat kesulitan butir, khususnya dengan menambahkan item pada level kesulitan rendah hingga menengah agar asesmen lebih sensitif terhadap variasi kemampuan mahasiswa semester awal.

Penelitian selanjutnya disarankan melibatkan sampel yang lebih besar dan heterogen serta memperluas sebaran tingkat kesulitan item. Selain itu, integrasi analisis Rasch dengan asesmen formatif perlu dikaji untuk memperkuat peran asesmen dalam perbaikan pembelajaran pendidikan teknik.

5. PERNYATAAN PENULIS

Penulis menyatakan bahwa tidak terdapat konflik kepentingan terkait penerbitan artikel ini. Penulis menegaskan bahwa naskah artikel bebas dari plagiarisme.

6. REFERENSI

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*.
- American Educational Research Association (AERA), American Psychological Association (APA), & (NCME), N. C. on M. in E. (2014). *for Educational and Psychological Testing* (Vol. 0000000000).

- Bond, T., & Fox, C. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, 3rd ed. Routledge.
- Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th Edition). Routledge. <https://doi.org/https://doi.org/10.4324/9780429030499>
- Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE—Life Sciences Education*, 15(4), rm4. <https://doi.org/10.1187/cbe.16-04-0148>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412.
- DeVellis, R. F. (2017). *Scale Development Theory and Applications* (Fourth Edition). SAGE Publication, 4, 256.
- Dong, Y., Xu, W., Huang, J., & Yann, K. (2025). Validating and refining a multi-dimensional scale for measuring AI literacy in education using the Rasch Model. *Humanities and Social Sciences Communications*, 12(1), 1–13.
- Feng, Y., Qiao, Y., Zhao, X., & Li, J. (2019). Study on Sample Size of Candidates Oriented to Online Test. *2019 14th International Conference on Computer Science & Education (ICCSE)*, 1006–1010.
- Fischer, G. H. (2006). Rasch modeling. In *Encyclopedia of Statistics in Behavioral Science: Everitt Behavioral* (pp. 1–9). Wiley. <https://doi.org/10.1002/0470013192.bsa540>
- Fitrah, M., Sofroniou, A., & Judijanto, L. (2024). Reliability and Separation Index Analysis of Mathematics Questions Integrated with the Cultural Architecture Framework Using the Rasch Model. *Journal of Education and E-Learning Research*, 11(3), 499–509.
- Hsu, J. L., Sung, R.-J., Swarat, S. L., Gore, A. J., Kim, S., & Lo, S. M. (2024). Variations in Student Approaches to Problem Solving in Undergraduate Biology Education. *CBE—Life Sciences Education*, 23(2), ar12. <https://doi.org/10.1187/cbe.23-02-0033>
- Jerosch-Herold, C., Chester, R., & Shepstone, L. (2017). Rasch model analysis gives new insights into the structural validity of the quickdash in patients with musculoskeletal shoulder pain. *Journal of Orthopaedic and Sports Physical Therapy*, 47(9), 664–672. <https://doi.org/10.2519/jospt.2017.7288>
- Kopp, J. P., & Jones, A. T. (2020). Impact of item parameter drift on Rasch scale stability in small samples over multiple administrations. *Applied Measurement in Education*, 33(1), 24–33.
- Linacre, J. M., George Engelhard, J., Wang, J., Boone, W., Haines, R. T., Royal, K., Marosszeky, N., & Haines, R. T. (2017). *Rasch Measurement Transactions*. 31(2).
- Maric, D., Fore, G. A., Nyarko, S. C., & Varma-Nelson, P. (2023). Measurement in STEM education research: a systematic literature review of trends in the psychometric evidence of scales. *International Journal of STEM Education*, 10(1), 39. <https://doi.org/10.1186/s40594-023-00430-x>
- Mayer, R. E. (2020). *Multimedia Learning* (3rd ed.). Cambridge University Press. [https://doi.org/DOI: 10.1017/9781316941355](https://doi.org/DOI:10.1017/9781316941355)
- O'Neill, T. R., Gregg, J. L., & Peabody, M. R. (2020). Effect of sample size on common item equating using the dichotomous Rasch model. *Applied Measurement in Education*, 33(1), 10–23.
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 1–14. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Saptadi, N. T. S., Septiani, S., Wardoyo, T. H., Hidayati, W., Thoif, M., Talindong, A., Audia, F., Nashiroh, P. K., Hadikusumo, R. A., & Suroya, L. F. (2025). *Pendidikan Vokasional:*

Teori dan Praktik. Sada Kurnia Pustaka.

- Siregar, T. (2025). Mengapa Validitas dan Reliabilitas Penting dalam L&D (Pembelajaran & Pengembangan): Apa yang Dapat Anda Lakukan Tentang Hal Ini. *JURNAL PEMIKIRAN DAN PENGEMBANGAN PEMBELAJARAN*, 7(1), 1–12.
- Tajuddin, N. I. I., Abas, U.-H., Aziz, K. A., Nor, R. N. H., Izni, N. A., Sudin, M. N., Anim, N. A. H. M., & Noor, N. M. (2025). Content Validity Assessment Using Aiken's V: Knowledge Integration Model for Blockchain in Higher Learning Institutions. *International Journal of Advanced Computer Science and Applications*.
- Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education*, 58(8), 922–933.
- Vincent, J. I., MacDermid, J. C., King, G. J. W., & Grewal, R. (2015). Rasch analysis of the patient rated elbow evaluation questionnaire. *Health and Quality of Life Outcomes*, 13(1). <https://doi.org/10.1186/s12955-015-0275-8>
- Wind, S. A., & Lugu, B. (2024). Combining nonparametric and parametric item response theory to explore data quality: Illustrations and a simulation study. *Applied Measurement in Education*, 37(2), 109–131.
- Xue, J., Chen, M., & Su, J. (2025). Evaluating the psychometric quality of classroom english assessments through the Rasch model. *Lecture Notes on Language and Literature*, 8: 6-13. <https://doi.org/http://dx.doi.org/10.23977/langl.2025.080502>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, 4(2), 165.